A Novel **Consensus Mapping** and **Enrichment Framework** for the Epigenomic Characterization of **TAD Boundaries**

[1]Krrish Anjan Ravuri

[1]Connor Frankston, [1,2]Dr. Gürkan Yardımcı

1Dept. of Biomedical Engineering, Cancer Early Detection Advanced Research, Knight Cancer Institute, Oregon Health & Science University.
2Division of Oncological Sciences, Cancer Early Detection Advanced Research, Knight Cancer Institute, Oregon Health & Science University.

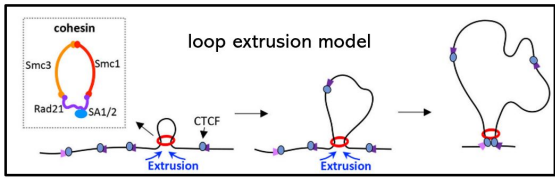
KNIGHT CANCER INSTITUTE
Oregon Health & Science University

Partnership for Scientific Inquiry
Building tomorrow's researchers today

Background

Topological Associated Domains (TAD) are neighborhoods of chromatin, (100kp - 2mb), where DNA sequences interact more frequently with each other than regions outside the domain. TADs play an important role in gene regulation and facilitating regulatory contacts between regulatory regions and their targets genes.

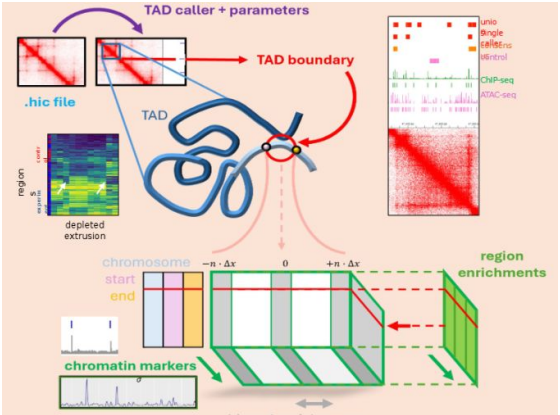
Loop extrusion is the prevalent hypothesis explaining how TADs form, yet other mechanisms, such as transcription related phase separation and compartmentalization, have been suggested.



Objective

The Objective of this project is to define a consensus TAD boundary algorithm which can identify high confidence TAD boundaries, and perform epigenomic characterization of said boundaries.

Graphical Abstract



Methodology

1. **Data source: Encode**
   - **Public Hi-C(GM12878, K562) - 10-25kb bins**
   - **Matching ChIp-seq, ATAC-seq & RNA-seq**
2. **Boundary Calling**

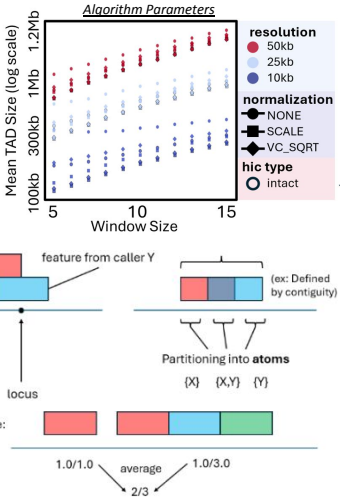
We used **Top Dom** with 20 different parameters sets, using a widely adopted insulation metric to identify TADs.

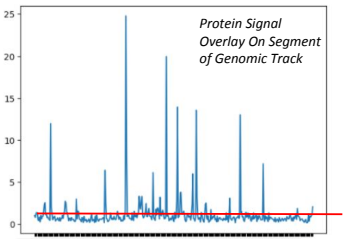
3. **Consensus Algorithm**

The consensus algorithm uses the common overlap of each of the TAD callers to find true consensus boundaries

4. **Test Correction:**

The algorithm utilizes multiple test correction supporting Bonferroni, Benjamin-Hochberg, and Sidak to correctly compare TAD calls with the null distribution





5. **Denoising:**

We modeled noise in flanks of genomic regions at TAD boundaries to clean & quantify ChIP-Seq signals of chromatin binding proteins

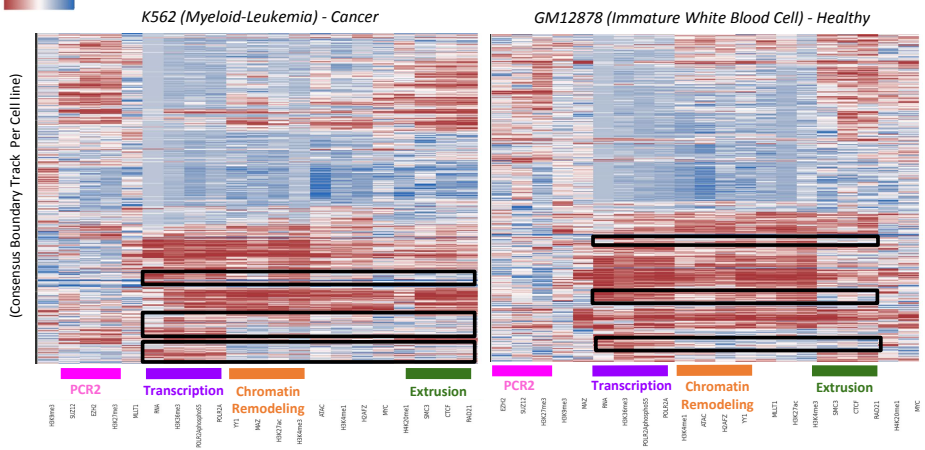
6. **Normalization:**

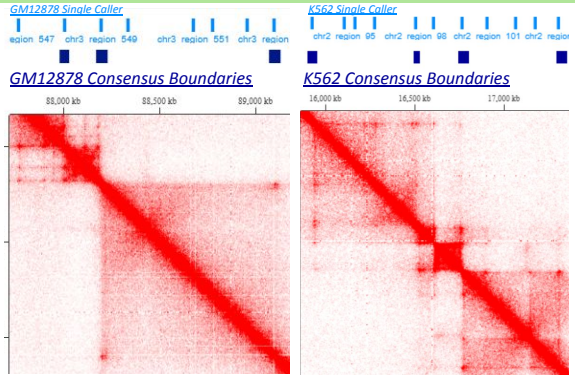
Heterogenous marker enrichments are normalized by a Yeo-Johnson Power Transform.

Results

Enriched, Normal, *Depleted*

Normalized Enrichment Heatmaps





Conclusion

- Consensus framework **reduces** caller/parameter **noise** and finds well-informed TAD boundaries
- Demonstrates **heterogeneous/different** mechanisms of TAD formation and its potential downstream role seen through different chromatin binding patterns at consensus tad boundary regions.

Next Steps

- **Scale-out:** apply to **larger cancer panel** (solid tumors, hematologic malignancies).
- **More features:** Include Loops, Stripes, Structural variations.
- **Algorithm refinement:** train machine-learning model on consensus labels to predict boundary type directly from sequence + epigenomics.

Acknowledgements

Internship: Partnership for Scientific Inquiry
Mentors: Dr. Galip Gurkan Yardimci, Connor Frankston.
Department: OHSU Knight Cancer Institute
Data sources: ENCODE
Software & Compute: OHSU Arc Cluster, JuiceBox